Technical Brief

New Methodology: a hybrid census to generate spatially-disaggregated population estimates

Publication Date: December 2017

Key words: Census, hybrid census, satelite imagery, geospatial data, Afghanistan

Population numbers at local levels are fundamental data for many development applications, including the delivery and planning of services, election preparation and response to disasters. In resource-poor settings recent and reliable demographic data at subnational scales can often be lacking or incomplete. National population and housing census data can be outdated, inaccurate or missing key groups or areas. Further, accurate boundary data are often limited, and high rates of migration and urban growth make existing data quickly outdated.

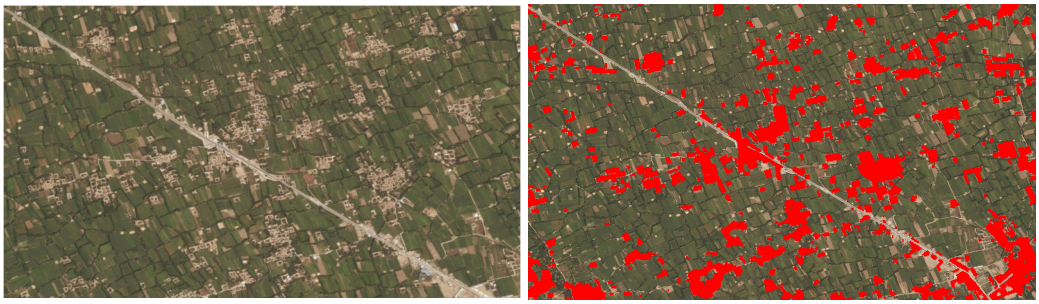
Example: The last population census in Afghanistan was conducted in 1979 and a new census is not foreseen until after 2020. UNFPA, Flowminder/WorldPop and the Afghanistan National Statistical Office have worked together to develop an innovative methodology to estimate subnational population counts. By combining survey, GIS and satellite data into a statistical model to predict numbers in areas where no population data were available, it was possible to generate high resolution maps of population estimates disaggregated by age and sex, together with uncertainty metrics. The data will be used to support decision making and the measurement of the 98 SDG indicators that require population data. The collaboration has also contributed to further strengthening capacities of the Central Statistics Office and other related government agencies to generate population estimates based on the integration of geospatial data and ground surveys.

Within the 2030 agenda, the growing requirement for spatially disaggregated population data has triggered the exploration of new data sources at different geographical scales and time periods, especially in highly stressed countries and countries without a recent population census. Advances in the availability of detailed satellite imagery, geo-positioning tools for field surveys, statistical methods and computational power are enabling the development and application of approaches

that can estimate population distributions at fine spatial scales across entire countries, in cases where population and housing census data cannot be conducted.

Figure 1 shows how buildings and settlements can be automatically detected from detailed satellite imagery to provide valuable geospatial data to guide population estimation and mapping.

Figure 1 (left) High resolution satellite image of a rural area of Afghanistan; (right) Settlement areas detected automatically by computer algorithms.



The goal of a hybrid census is to produce population estimates for small areas or uniform, detailed grids in the absence of a traditional national census. Hybrid censuses rely on complete counts of population within small, defined areas, through 'micro-census surveys', selected across an area of interest, and collected relatively rapidly and at a fraction of the cost of a full national census. Statistical models are then used to link these micro-census data to spatial data with full coverage over the regions of interest to predict population numbers in the unsampled locations. By aggregating these high-resolution predictions, population totals can also be produced for administrative units or for the national level if required.

Figure 2. Schematic of a hybrid census

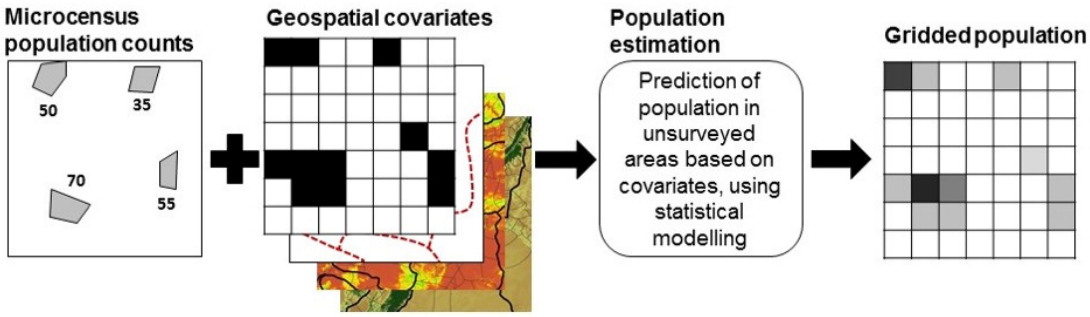
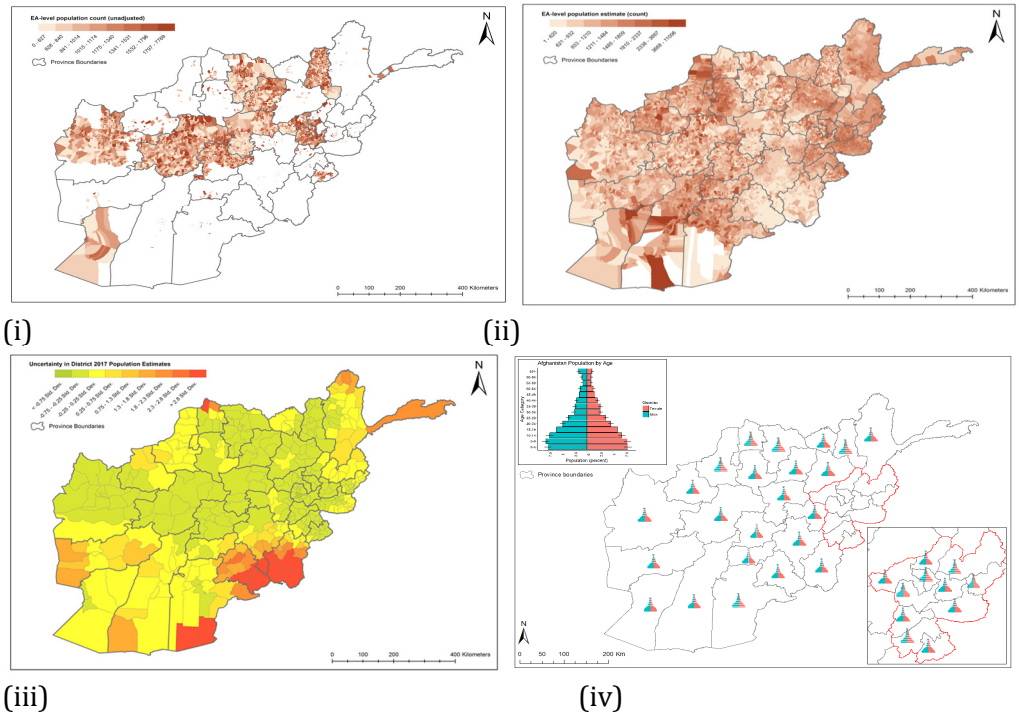


Figure 2 shows the different layers and steps involved in a hybrid census: i) micro-census population enumeration; ii) linkage with relevant geospatial datasets; and iii) definition of the functional relationships between geospatial datasets and the population data to predict populations in un-surveyed locations.

Population data for a sample of areas across the area or country of interest are needed as a primary input for a hybrid census. These data may come from a partial census, census-like population survey or a specifically designed micro-census survey. A requirements for these data are robust geo-referencing of the geographical areas where population data has been acquired, and a design that captures as best as possible the range of densities, demographics and environmental dimensions that exist across the area of interest. Population enumeration can be conducted within administrative units or within other arbitrarily designed polygons, as long as the population data is explicitly linked to the correct geographical area. Global Positioning Systems (GPS) can be used to ensure enumeration is occurring in the correct location, and the inclusion of GPS technology in smart phones and tablet computers can integrate navigation and the recording of geographical coordinates into enumeration activities, minimising locational error and human effort.

Figure 3. (i) available recent survey and census data in Afghanistan; (ii) Fully predicted population total estimates for Afghanistan in 2017; (iii) Standard deviation map showing levels of uncertainty in predicted population numbers; (iv) Estimated population pyramids for 2017.



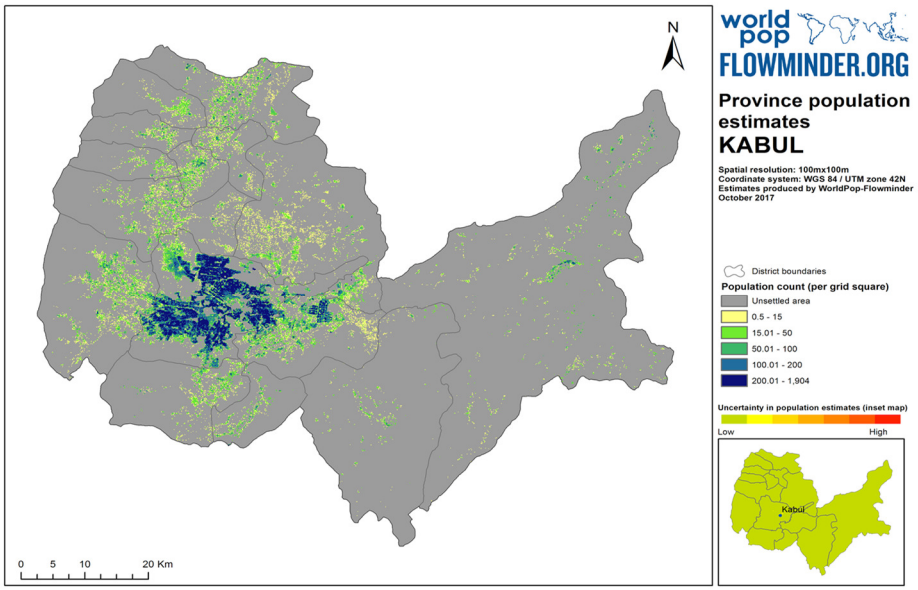
(i) (ii)

(iii) (iv)

Figures 3 i to iv show the outputs from the Afghanistan analyses, combining the enumeration area counts with geospatial covariates in a spatial statistical model to predict population numbers and age/sex structures into all unsampled areas, with uncertainty measures provided. At the district and province levels, accuracy assessments measuring observed vs predicted population numbers showed correlation coefficients of >0.95. In the case of Afghanistan, the predictions were disaggregated to 100x100m grid squares to provide a consistent and flexible representation for integration with other datasets (e.g. calculating numbers of people within a certain travel time of a health facility) and summarization to different administrative units. Figure 4 shows the gridded output for Kabul Province. Hybrid census approaches can be produced through direct estimation at the grid square level also, as has been undertaken in Nigeria for support in polio vaccination planning and tracking.

Figure 4. Gridded 100x100m estimates of population counts for Kabul province, 2017,

with uncertainty map shown in inset box.



The covariates used for hybrid censuses should be i) strongly correlated to population density and ii) available consistently across all areas where the population estimation is required. Population estimation approaches have previously been categorised as utilising the relationships between population and covariate data representing the following: 1) built up areas, 2) areas of specific land use types, 3) counts of dwelling units, 4) satellite derived measures such as spectral radiance; or 5) socio-economic or physical characteristics. Although access to high quality, spatially comprehensive datasets representing some of these characteristics has traditionally been difficult in resource-poor settings, advances in image-processing techniques, computational power, and the increasing availability of very high resolution satellite imagery means that the production of high quality covariates for many settings is increasingly feasible.

Several limitations to the approach need to be considered: Micro-census data must be collected with the same care and rigor as a full census. Further, population data are highly political and contentious, affecting all per capita rate estimates, shifting political representation and changing claims to power or resources. Conflicts, environmental hazards, or large population displacements make it difficult to accurately collect information on the dynamic population, and misrepresenting populations at risk

should be avoided.

This approach can never replace the rich production of data on the individual, family, household or community generated by a traditional population and housing census. However, where a traditional census cannot be fully executed in all locations of a given country due to insecurity or other concerns, then this hybrid approach could produce population estimates for small areas or uniform, detailed grids in the absence of traditional census data.

This technical brief has been prepared jointly by Sabrina Juran, Technical Specialist, Data and Population Analysis, Population and Development Branch, Technical Division in collaboration with Andrew Tatem, Professor of Spatial Demography at the University of Southampton, and Director of WorldPop/Flowminder.